Micro-CT for Saw Mark Analysis on Human Bone

Norman, D.G.,[1] Baier, W.,[1] Watson, D.G.,[2] Burnett, B.,[3] Painter, M.,[4] & Williams, M.A.[1]*

Danielle G Norman, MEng (d.norman.3@warwick.ac.uk)

Waltraud Baier, MA, MSc (v.baier@warwick.ac.uk)

Derrick G Watson, BSs, MSc, PhD, DSc (d.g.watson@warwick.ac.uk)

Brian A Burnett, DFD MBIFD DAPT MRSPH (brian.burnett@uhcw.nhs.uk)

Michelle Painter, MA (m.painter@west-midlands.pnn.police.uk)

Mark A Williams, BEng, MSc, PhD, CEng, FIMechE (m.a.williams.1@warwick.ac.uk)

[1] WMG, International Manufacturing Centre, University of Warwick Coventry, CV4 7AL, UK

[2] Department of Psychology, University of Warwick Coventry, CV4 7AL, UK

[3] Anatomy & Surgical Training Centre Manager, UHCW NHS Trust, Coventry, CV2 2DX, UK

[4] Assistant Director of Forensic Services, West Midlands Police, Birmingham, B4 6NQ, UK

*Corresponding author

FUNDING - This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENTS - We would like to thank Samual Horne, Davina Cowie and Leigh Nash for their excellent assistance in the lab as well as volunteering to create the toolmarks.

Micro-CT for Saw Mark Analysis on Human Bone

KEY FINDINGS

- Micro-CT allows visualisation and reliable measurement of saw mark properties.
- Different tools create significantly different toolmarks on human bone.
- Tissue presence and saw action significantly and inconsistently affect toolmarks.
- A toolmark width regression model from Study 2 predicted tool widths in Study 1.

ABSTRACT

In toolmark analysis, microscopy techniques, such as micro-CT, are used to visualise and measure toolmarks left on bones by a tool. In dismemberment cases, properties such as the width of the saw mark can provide cues to which tool was used by the culprit. The aim of the current study was to establish whether; i) micro-CT is an appropriate imaging technique for saw mark analysis, ii) toolmarks statistically differ when created with different tools, iii) toolmark width can predict tool blade width, and iv) toolmarks differ if created under different methodological conditions. Across two experiments, 270 saw marks were created using eight tools with either a controlled or free saw action on either fleshed or defleshed human long bone. Toolmarks were micro-CT scanned and seven toolmark properties were categorised or measured by two independent raters. The current study found that; i) micro-CT was found to be a powerful and reliable imaging method for the visualisation and measurement of saw mark properties, ii) toolmark properties differed significantly within and between various methodological conditions ($p<.001$) when created by eight different tools, iii) a regression model developed using toolmark widths from Experiment 2 overall predicted 94% of tool widths in Experiment 1, and iv) methodological factors such as tissue presence and saw action significantly and inconsistently influenced toolmark properties for different tools. The study further validates the use of mirco-CT for saw mark analysis and demonstrates the potential of using toolmark properties to determine the tool used in cases of dismemberment. Given the effects that methodological factors such as tissue presence can have on toolmark properties, future studies should use experimental set ups with fleshed human tissue and use a free saw action

Keywords: Micro-Computed Tomography (Micro-CT); Forensic Toolmarks; Dismemberment; Saw; Bone

INTRODUCTION

1.1.Toolmark Analysis

In toolmark analysis, microscopy techniques are used to measure toolmark properties on defleshed bone surfaces. Previous studies have used a wide range of experimental approaches for the study of saw marks [1-18]. The differences between these studies include; i) tissue types (human, animal and synthetic analog), ii) and whether the bone was fleshed, semi-fleshed or defleshed, iii) sawing actions (controlled actions such as using miter saw, free saw actions using an unrestricted human volunteer), iv) number of volunteers, v) range of saws tested both within and between class, vi) number of toolmarks created, vii) imaging methods applied (Digital microscopy, Scanning Electron Microscopy, Epifluorescence, stereomicroscopy, radiography), and viii) the method of analysis (e.g., correlation, regression, decision trees, classification trees and random forest classifiers).

1.2.Methodological Issues

It is likely that these methodologies make a significant difference to the toolmarks created and the information obtainable when using different imaging methods. For example, Nogueria et al., found significant differences between the toolmark widths, created with 4 saws, on pig (50 toolmarks) and human (120 toolmarks) femurs suggesting that animal substitutes, although practical, are not adequate substitutes for human tissue [13]. Further research on other methodological variations is therefore important, however, to the authors' knowledge, this has yet to be conducted.

1.3. Micro-CT Imaging

Most toolmark analysis studies use microscopy methods to allow visualisation and measurement of toolmark properties. However, these methods; i) require that the tissue be defleshed so that the toolmarks are visible, ii) do not allow inspection of the internal or obscured parts of a toolmark, and iii) rarely allow 3-dimensional reconstruction of the toolmark. Micro-Computed Tomography (micro-CT) is a form of radiography that, unlike medical grade CT, allows for high resolution imaging (<100µm) of toolmarks [14-18]. As well as the practical benefits, such as non-destructive imaging, Norman et al., found that micro-CT could reveal additional toolmark properties not previously possible with microscopy techniques when imaging knife marks [19]. In addition, the visualisation capabilities of micro-CT have been demonstrated in a recent dismemberment case [16]. Nevertheless, micro-CT is a relatively new imaging method and its use in saw mark analysis has only recently been explored. Two recent studies by Pelletti et al., highlighted the visualisation advantages of micro-CT [17] as well as testing its accuracy and reliability for toolmark measurements [18]. However, both of these initial studies recognised their small samples sizes of toolmarks (24 and 32 respectively) and saws (4 and 3 respectively) and therefore Pelletti et al., encourage further work.

1.4. Current Study

The overarching focus of this research was to analyse toolmarks created under different methodological conditions and imaged using micro-CT to expand the previous saw mark literature. Our research aimed to establish whether; i) Micro-CT is an appropriate technology for visualising and measuring qualitative and quantitative toolmark properties and if so, what is the reliability of these measurements when extracted from micro-CT data, ii) Toolmark properties differ statistically when created with different tools, iii) Across all conditions, there is a significant correlation between toolmark width and tool blade width and if so what percentage of toolmark width can predict tool blade width, and iv) Toolmarks differ if created under different methodological conditions namely fleshed vs. defleshed bone with either use of a controlled vs free sawing technique.

MATERIALS AND METHODS

2.1. Tools Sourced

Eight newly purchased tools belonging to separate saw and knife classes were selected to represent those that are commercially available and commonly found in households and therefore likely to be used in cases of dismemberment [9, 20-21] (Fig.1.). The tools consisted of two electric power saws, four hand powered saws, and two knives, one of which was serrated – tool properties are provided in Table 1. Note that tool blade width was measured by taking the mean of 30 measurements along the length of the blade using dialed calipers.

[Figure 1]

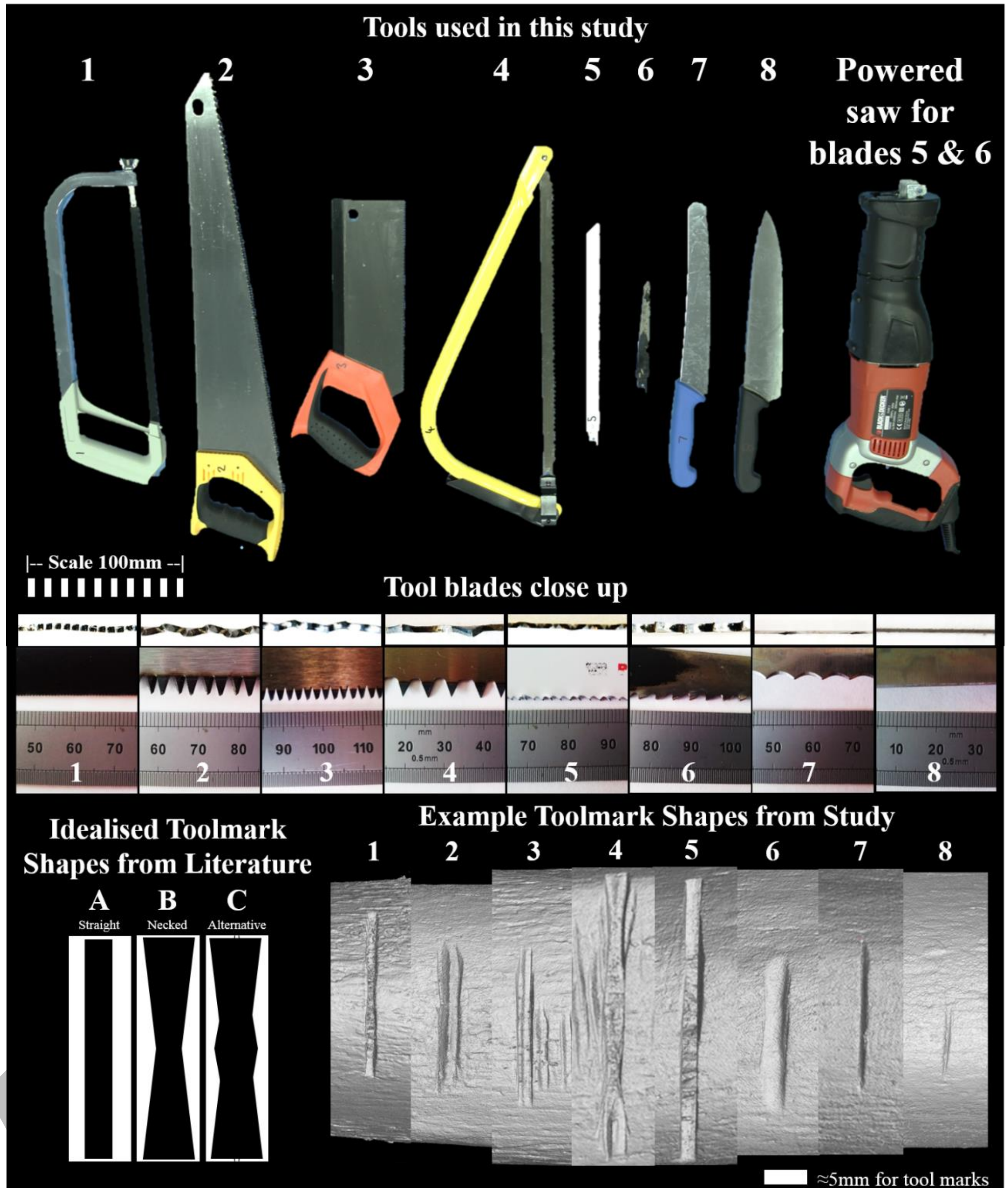

Figure 1: The eight saws used in this study and respective blade profiles and toolmark edge shapes described in the literature (A-C) with corresponding example toolmark edges observed in this study (1-8) with their respective tools below. Toolmark: A) straight edge shape with near parallel edges seen in toolmarks 2-3,5-8 and typical of raker set blade but also seen with wavy and alternative sets; B) necked edge shape with a distinct necking in around the centre seen with toolmark 1 and typical of wavy set blades; C) alternative edge shape with both narrow and wide aspects seen in toolmarks 4 and typical of alternating set blades.

[Table 1]

Table 1. Properties of the tools used in this experiment

Saw	Saw Type	Set	Set Width (mm)	Teeth per Inch (TPI)	Tooth Type	Length (mm)	Width (mm)
1	Hacksaw	Wavy	0.7	32	Rip cut	300	0.69 ± 0.04
2	Hand	Alternating	2.0	7	Crosscut	550	1.41 ± 0.10
3	Tenon	Alternating	1.1	11	Crosscut	250	0.75 ± 0.04
4	Bow	Alternating	1.3	5[1]	Crosscut	375	0.78 ± 0.11
5	Sabre	Raker	1.4	12[2]	Rip cut	300	1.10 ± 0.10
6	Reciprocating	Raker	1.6	6	Rip cut	130	1.47 ± 0.11
7	Carving Knife	-	-	4	-	24	0.46 ± 0.07
8	Cook's Knife	-	-	-	-	24	0.52 ± 0.08

[1]The bow saw is alternating set with teeth in groups of threes, with larger distance between them (7:20 ± 0:02) mm compared to (5:28 ± 0:02) mm.
[2]The sabre saw does not have a well-defined TPI and varies from 10 to 14 TPI with a periodicity of approximately 40mm.

2.2. Toolmark Creation

2.2.1 Experiment 1

Eight cadaveric human legs (from the hip joint down) were sourced following full ethical approval by the author's institution and following standard Human Tissue Authority guidelines. Human femur and tibia were chosen because of the prevalence of dismemberment cases that involve bisection of the long bones [20]. Toolmarks were created on four human donor's left and right leg consisting of a femur and tibia bone and sawn on both the anterior and posterior sides. Two sawing actions were used; 'controlled' and 'free'. To create controlled marks, 15 straight saw strokes, push and pull, with moderate force were applied for all toolmarks placed approximately 2cm apart. In the free condition the volunteer was given no restrictions except to create a false start toolmark – the aim here was to closely simulate a dismemberment thereby creating real-world toolmarks. Half of the toolmarks were made on fully fleshed bone (Fig.2A.) and half on defleshed bone (Fig.2B.). Half of the toolmarks were created on the bone moving from inferior to superior with the other half being reversed. All the above factors were fully counterbalanced with the order of the eight saws randomised for each individual bone side. This design resulted in 248 saw marks across 2 donors/4 legs/8 bones/16 bone sides (see Supplementary Materials - Table 1). Although each tool was used 30 times, any wear of the blade was considered irrelevant based on work by Freas et al., showing that the effects of saw wear do not significantly hinder examination of class characteristics [9]. Following toolmark creation the bones were manually defleshed, individually packaged and refrigerated prior to micro-CT imaging.

2.2.2 Experiment 2

In order to explore the potential of statistically predicting an unknown tool from a toolmark, a separate experiment was conducted following Experiment 1 to develop an independent regression model for application to the toolmarks in Experiment 2. Therefore, in Experiment 2, four cadaveric human tibiae were sourced in the same manner as Experiment 1 with the exception that only partially defleshed tibiae were available. Each tibia was fully defleshed and set in cylindrical mould with a 10% mixture of ballistics gelatin. Ballistics gel is a material that is often used to simulate the properties of human soft tissue. The aim of setting the bones in this material was to provide a more realistic power transfer from the saw in comparison to direct clamping of the bone in a vice which can add compressive and tensile forces to the bone. The moulds were approximately the diameter of a human leg (Fig.2D.). Five false start marks per saw were created by an adult male volunteer. Therefore, across the five tibiae, 40 controlled saw marks were created. Following toolmark creation the tibiae were dissected from the ballistic gel (Fig.2E.) and then individually packaged and refrigerated and micro-CT imaged in the same way as Experiment 1.

2.3. Micro-CT Imaging

A Nikon XT H 225/320 LC micro-CT scanner was used to image all toolmarks. Scan parameters were 200kV, 30W, no filter and 1.4s exposure which resulted in resolutions of approximately 50µm with total scan time of approximately 74 minutes. 3-dimensional micro-CT data were reconstructed using Nikon's Proprietary software, *CT Pro* and then exported to VGStudio Max (examples in Fig.2C. and Fig.2F.) with each micro-CT scan being calibrated using an artefact with known absolute dimensions [22-23].

[Figure 2]

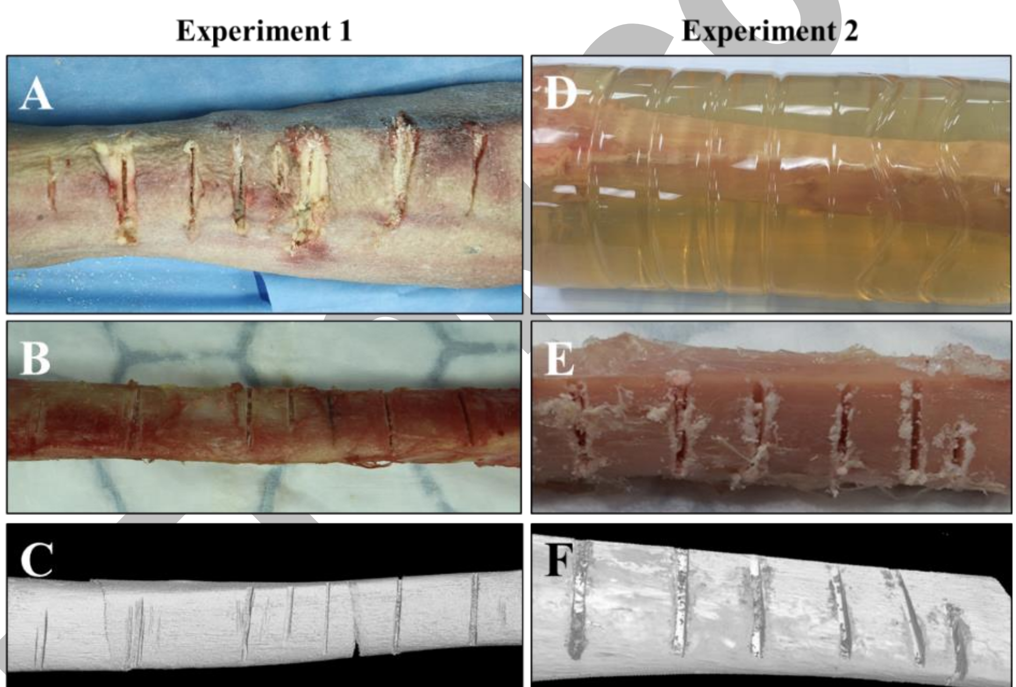

Figure 2. Experiment 1: A) Fleshed human femur with saw marks; B) Defleshed saw marked bone; C&F) Micro-CT model of the bone in B&E respectively. Experiment 2: D) Human femur set in ballistics gelatin; E) Saw marked bone.

2.4. Toolmark Analysis

2.4.1. Image Extraction

Two images per toolmark were exported from the micro-CT data to allow the toolmark properties to be measured. A 3D rendered top-down image was acquired to determine the toolmark shape and a 2D cross-section of the toolmark to determine the toolmark profile and any quantitative measurements. The 3D top-down image was captured by first rotating the micro-CT 3D model so that the full length of the longest mark (if there were multiple marks from that trial) was visible; fitting a best-fit 2D reference plane to the toolmark floor where visible; Aligning the 3D perspective view so that the plane was flat thereby resulting in a birds-eye view of the toolmark; finally, a high resolution scaled image of that view was exported. The toolmark profiles were taken from a cross-sectional view, perpendicular to the reference plane, half way along the length of the saw mark or until a complete floor was visible when moving from the mid-length position towards the saw entrance. If the edge shape was alternating, then the first necking position from the mid-length position was taken. If multiple marks were made, then the deepest and/or most complete saw mark was used as the profile shape. Furthermore, if the shape was incomplete due to the saw fully penetrating the cortical bone (thereby removing the floor of the cut mark and exposing the trabecular bone) then the shape profile was taken at the first point between the mid-length and the toolmark entrance where a complete floor was apparent. The two images for each toolmark were exported with parameters set to ensure scaling consistency between the images. To achieve this in VGStudio Max the 'zoom factor' and 'relative scale bar size' were set to approximately 800% and 25% respectively with small adjustments required for each bone sample to ensure that all scale bars were equal in pixel size and representing 3mm in absolute size. All 2D images were exported at max resolution as jpeg files with a resolution 2138x1273. Other factors set were: Ruler Sections – 3; Cell Boarders – 20 pixels from the bottom of the image; Scale Font – Times New Roman, 36pt; Scale DPI – 200%. The 3D top-down images and the profile images were then imported into ImageJ, scaled, and then used to categorise the toolmark shapes and profiles and measure the toolmark properties.

2.4.2. Toolmark Measurements

Toolmark properties were measured using a similar approach outlined in previous work by Norman et al., [19]. Seven observations for each saw mark were measured or categorised by two, independent blind raters with previous toolmark research and casework experience and with access to only the methodology section of this paper particularly Fig.1. and Fig.3. There were two categorical classifications: 1) 'Edge shape' - straight, necked or alternating [10], Fig.1. shows idealised examples and examples from this study with corresponding tools; 2) 'Profile Shape' – either concave shaped; round, square, V and Y or convex shaped; W and W-truncated (Fig.3). shows an example for each tool; Up to five quantitative properties were measured with two for all toolmarks: 4) minimum toolmark width at floor and 5) wall angle, and three additional quantitative measures were taken for convex profiles: 6) trough height, 7) trough angle deep, and 8) trough angle shallow. Prior to these measurements it was agreed that rater 1's observations would be used for all statistical analyses. Statistical agreement for categorical ratings would be calculated using Cohen's kappa [24] for Experiment 1 and Experiment 2 separately and for each independent condition in Experiment 1 i.e. controlled and free saw actions on both fleshed and defleshed tissue. Agreement for quantitative measures would be calculated for the same groups but using the Interclass Correlation Coefficient (ICC) [25].

[Figure 3]

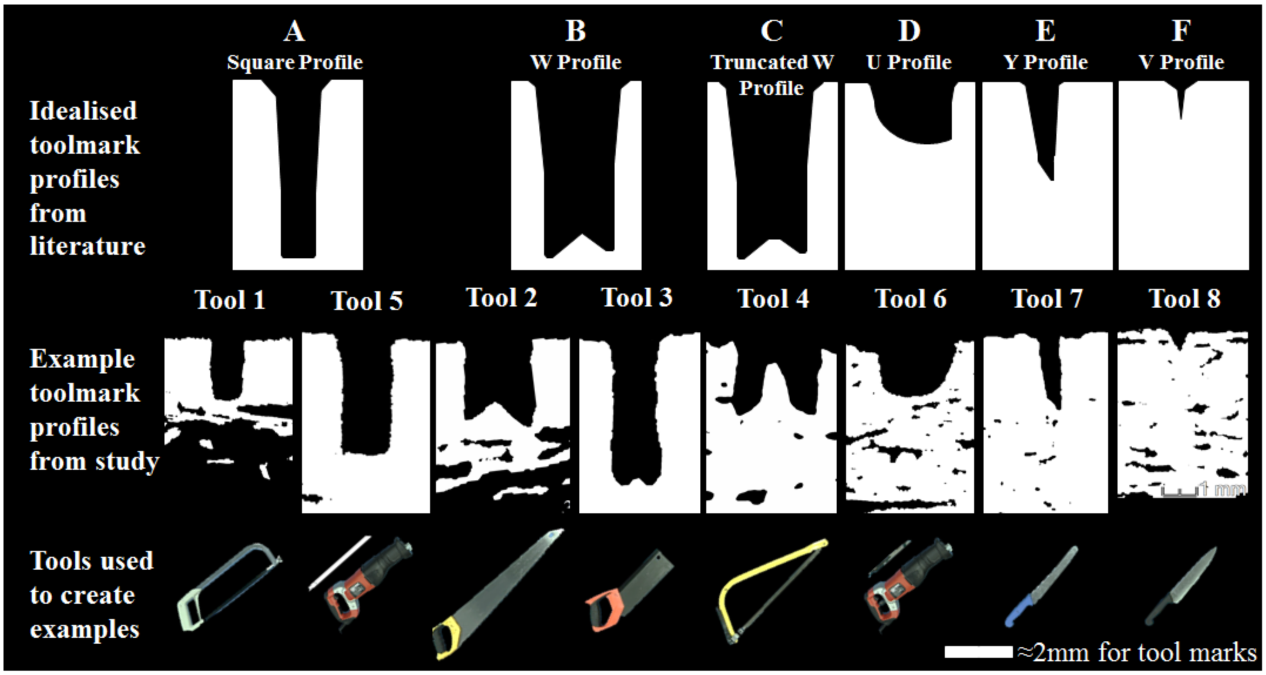

Figure 3. The top row shows examples of toolmark profiles both observed in this study and documented in the literature for saws and knives. These profiles include: A) Square profiles with straight near parallel edges B) W profiles with sharp trough peaks; C) Truncated W profiles similar to W profiles but with truncated trough peaks; D) U profiles which have a distinct oblong component; E) Y profile typically caused by serrate knives; and F) V profile typically seen with non-serrated knives. The middle row shows examples of these toolmark profiles created from each of the tools used this this study. The bottom row shows the corresponding tool used to create the toolmark seen in the middle row.

[Figure 4]

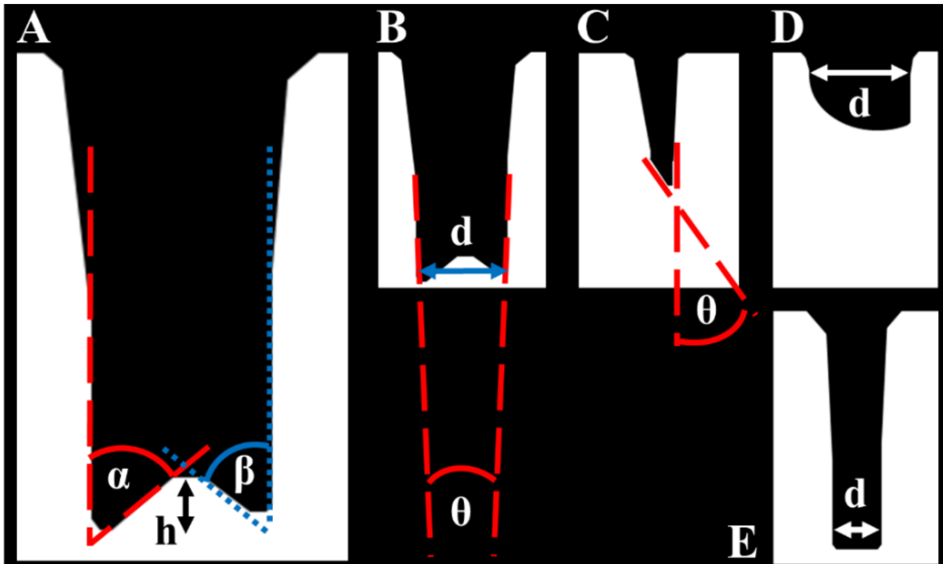

Figure 4. Examples of individual toolmark properties measured in this study. Trough Height h is the perpendicular height between the trough peak and a line drawn between the two trough mark floors. Trough Angle Deep α, is the angle between the trough peak and the adjacent wall for the deepest trough and similarly Trough Angle Shallow β, is the equivalent for the smallest trough. B) Wall Angle θ, is the angle between the two walls (this can be small for near parallel walls). Min Toolmark Width d is the minimum width between the two wall faces nearest to the toolmark floor (for: square profiles this is parallel to the floor); W profiles this is the distance between the ends of the trough marks; U, V and Y shaped profiles this is the width (parallel to the bone surface) between the two wall faces nearest to the surface of the toolmark. C) example of a θ angle for a Y profile. D) example of the width measure for a U profile. E) example of the Min Toolmark Width measurements for square profiles.

RESULTS

3.1. Micro-CT Imaging

Micro-CT was shown to be an appropriate technology for visualising and measuring qualitative and quantitative toolmark properties (Fig.5. and Fig.6.). Furthermore, measurements extracted from micro-CT data were shown to have a high inter-rator reliability. The inter-rater reliability of all toolmark measurements, 3780 in total across two raters, was compared using either kappa or ICC inter-rater statistical tests. With the exception of edge shape and trough shallow angle, agreement was either 'good' or 'excellent' across all conditions in Experiment 1 and Experiment 2. Measures of floor width in particular were highly consistent between the two raters. (Supplementary Materials – Table 2)

[Figure 5]

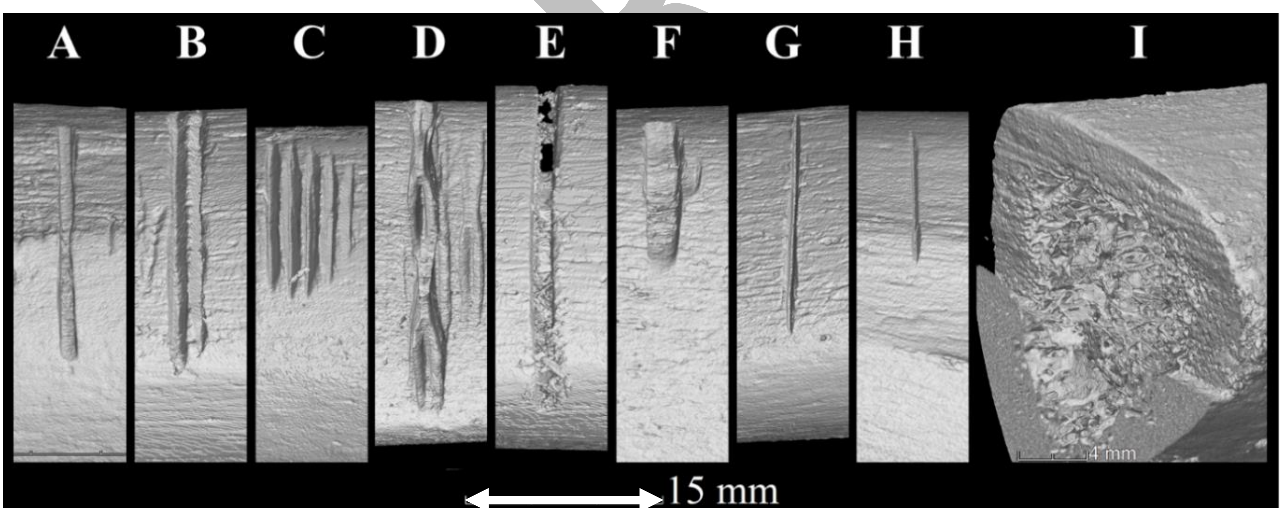

Figure 5. Examples of various qualitative toolmark properties including: A) Necked edge shape from tool 1; B) Full trough from a crosscut blade, tool 2; C) Multiple marks created by tool 3 due to tooth hop; D) Bone islands created by tool 4; E) Exit chipping and removal of the cortical bone from tool 5; F) U shaped toolmark from tool 6 showing floor dip 6; G) Y shaped knife mark from tool 7; H) V shaped knife mark from tool 8; and I) Striation marks from tool 5.

[Figure 6]

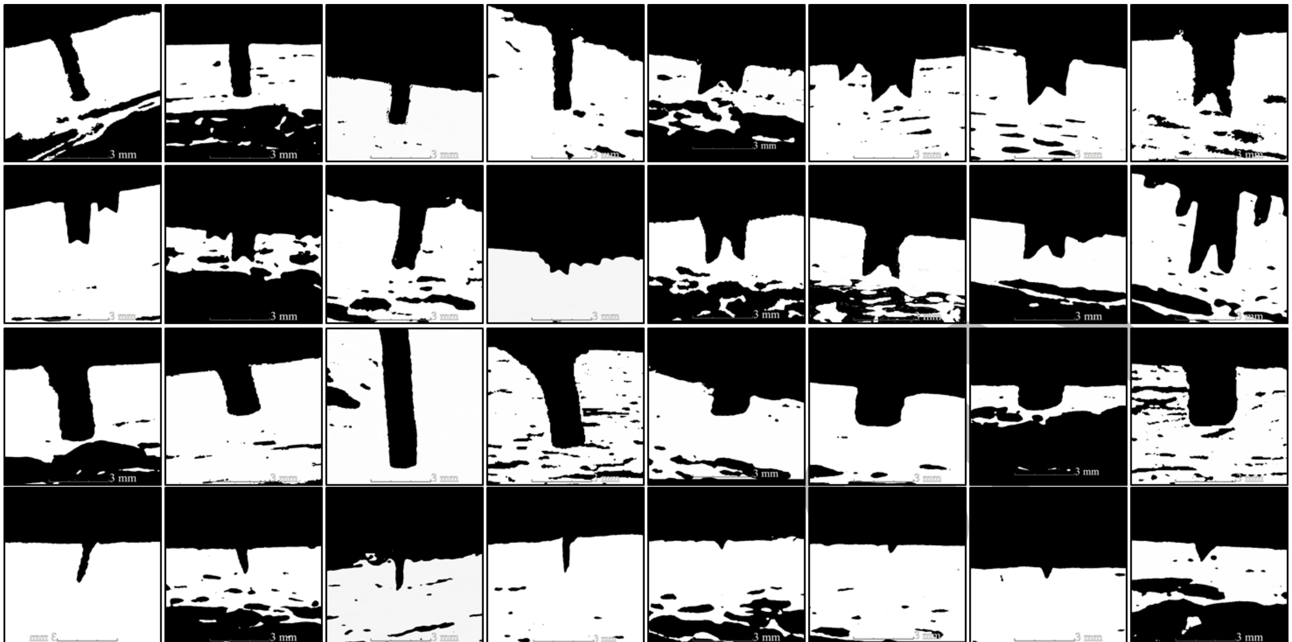

Figure 6. Four example toolmark profiles for tools 1-8 (running left to right) randomly selected from Experiment 1 and scaled.

3.2. Toolmark Differences

The results indicated that toolmark properties, both categorical and quantitative, differed significantly when created using different tools (Supplementary Materials – Table 3 and 4). The quantitative toolmark differences are presented in Fig.7. One-way Welch ANOVAs, for each quantitative toolmark property, revealed a significant difference between toolmarks for tools 1-8, ($p<.001$ except for Trough Shallow Angle which was $p=.004$). Minimum toolmark width appeared to be the most diagnostic property with follow up post-hoc tests revealing significant difference between all tools ($p<0.001$) except 2&6, and 7&8.

[Figure 7]

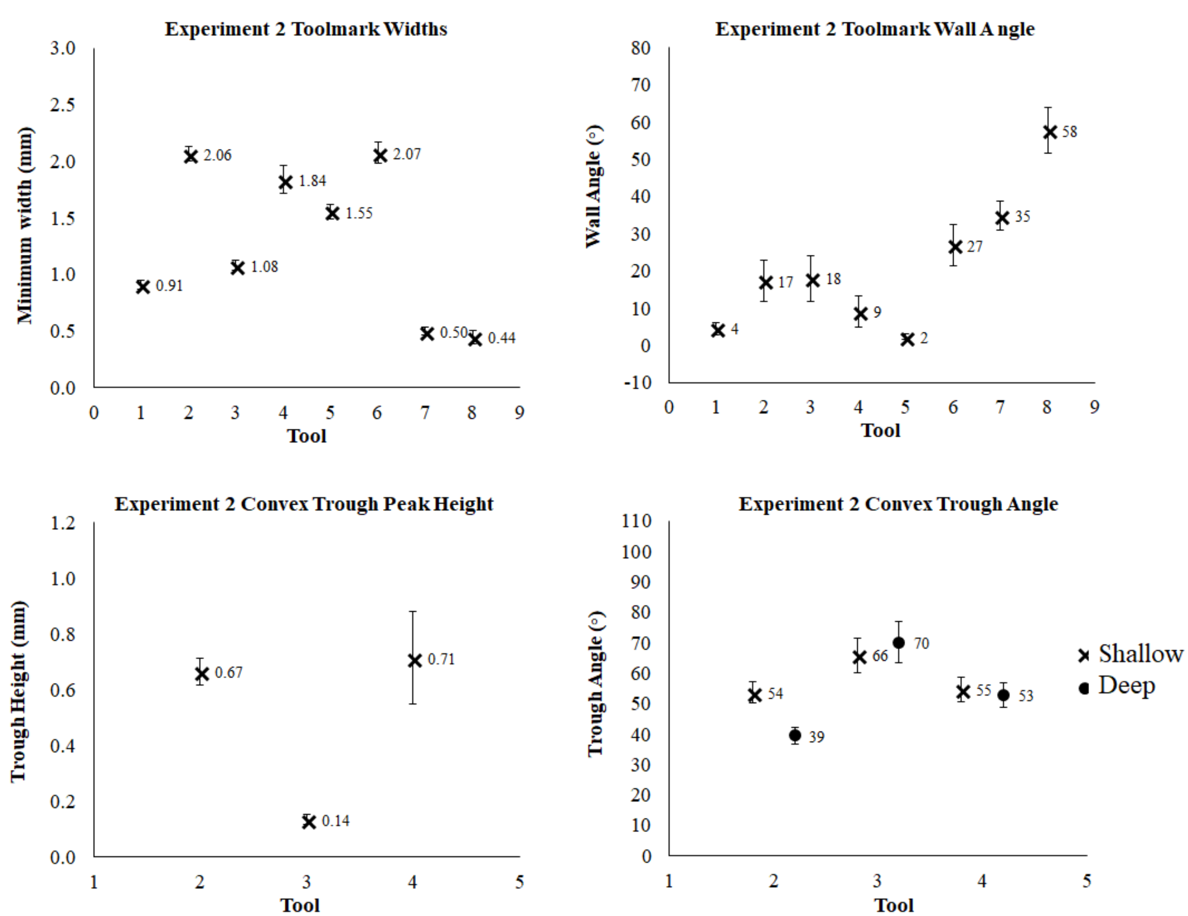

Figure 7. Mean and 95% confidence intervals for quantitative toolmark properties in Experiment 2 for each tool

3.3. Tool Prediction

Across all conditions, there was a significant correlation between toolmark width and tool blade width. Recall that in order to explore the potential of statistically predicting an unknown tool from a toolmark, a separate experiment was conducted following Experiment 1 to develop an independent regression model for application to the toolmarks in Experiment 2. In Experiment 2 minimum toolmark width was highly correlated with tool blade width, $r(40)=0.864$, $p<.001$. For Experiment 2 a linear regression with predictor variable of toolmark width and outcome variable of tool blade width was conducted. Assumptions: Linearity was established by visual inspection of a scatterplot. There was independence of residuals, as assessed by a Durbin-Watson statistic. There was homoscedasticity, as assessed by visual inspection of a plot of standardised residuals versus standardised predicted values. Residuals were normally distributed as assessed by visual inspection of a normal probability plot. Toolmark width accounted for 71% of the variation in tool blade width with adjusted $R^2 = 71\%$, a large size effect according to Cohen (1988). In Experiment 2 toolmark width significantly predicted tool blade width, $F(1, 39) = 94$, $p < .001$. The prediction equation was: Toolmark Width (mm) x 0.51 + 0.23 = Tool Blade Width (mm). At 95% confidence this linear model correctly predicted 95% of tool blade widths from toolmark widths (Fig.8.). (Note in Experiment 1 toolmark width significantly predicted tool blade width, $F(1, 237) = 581$, $p < .001$. The prediction equation was: Toolmark Width (mm) x 0.47 + 0.28 = Tool Blade Width (mm)). Applying the regression model from Experiment 2 to toolmark widths in Experiment 1 allowed the prediction of: 94% of tool blade widths across all conditions; 39% in the fleshed bone condition; and 19% in the fleshed, free saw action condition.

[Figure 8]

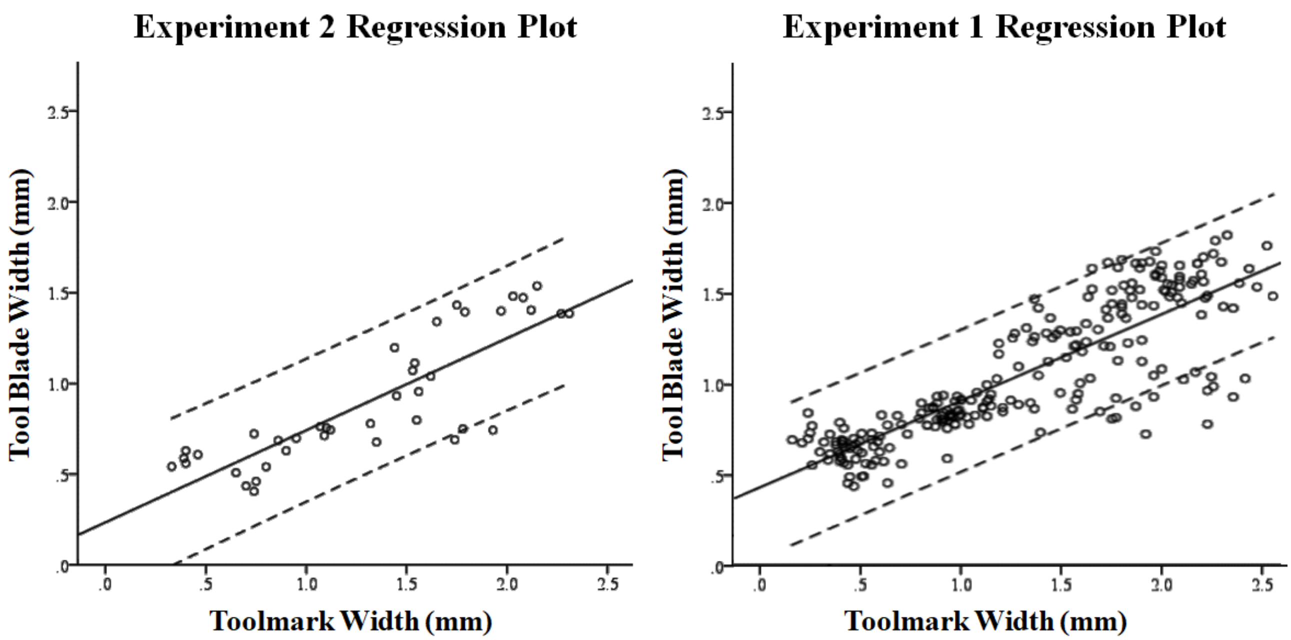

Figure 8. Scatterplot of tool blade width (mm) of tools 1-8 used in this study, against toolmark widths (mm) of tool mark created in Experiments 1 and 2. The solid line shows the linear regression fit with R^2 of 0.71 (for both graphs) and equations y = 0.51x + 0.23 (Left graph) and y = 0.47x + 0.28 (right graph). The dashed lines represent the individual confidence intervals at each prediction of the linear regression equation.

3.4. Methodological Factors

Toolmarks differed when created under different methodological conditions namely fleshed vs. defleshed bone and depended whether a controlled vs. free sawing technique had been used. The differences in toolmark width created by different tools is often studied on defleshed bone using a controlled saw action. To directly test whether toolmark widths in these conditions reflect what is more common in real forensic case, i.e. toolmark created on fleshed bone using a free sawing method, a three-way 2x2x8 ANOVA was conducted to determine the effects of tissue (fleshed & defleshed), saw action (controlled & free) and tool (1-8) on toolmark properties in this study. For toolmark width there was a statistically significant three-way interaction between tissue, saw action and tool, $F(1, 7) = 3.8$, $p = .001$. For the remaining quantitative toolmark properties (wall angle, trough angles and trough height) there was only significant main effect of tool ($p<.001$) therefore only toolmark width is explored here after.

Splitting the data by saw action. There was a statistically significant simple two-way interaction (2x two-way ANOVAs) between tissue and tool for both controlled saw actions, $F(1, 7) = 4.6$, $p < .001$, and for free saw actions, $F(1, 7) = 3.5$, $p = .002$. Further splitting by tissue there was a statistically significant simple main effect (4x one-way ANOVAs) of tool: for free saw actions on defleshed bone, $F(7, 19) = 149$, $p < .001$; for free saw actions on fleshed bone, $F(7, 23) = 110$, $p < .001$; for controlled saw actions on defleshed bone, $F(7, 23) = 246$, $p < .001$); and for controlled saw actions on fleshed bone, $F(7, 6021) = 254$, $p < .001$). All simple pairwise comparisons were run for all comparisons with Tukey adjustments applied. All pairwise comparisons were significant (p<.001) except for difference between tools 1&3, 2&6, 2&6, 4&6 and 7&8 for all simple main effects and tools 2&4 for free and controlled saw actions on defleshed bone only. These results indicate that in the majority of cases toolmark widths differed significantly between tools regardless of tissue presence or saw action. Fig.9. shows a graphical representation of these differences for each condition.

Splitting only by tool. Eight separate 2-way ANOVAs were conducted. For saws 1-3 and 7 there was no significant effect of saw action or tissue presence. However, for tool 4 there was a significant interaction between saw action and tissue, $F(1,25)=9.04$, $p=.006$ (Fig.9. – Tool 4). For tools 5 and 6 there was a significant effect for tissue presence, $F(1,31)=26.4$, $p<.001$ and $F(1,31)=21.5$, $p<.001$ respectively (Fig.9. – Tools 5 and 6). A further significant effect of saw action was found for tool 6, $F(1,31)=23.7$, $p<.001$ (Fig.9. – Tool 5) and tool 8, $F(1,30)=15.8$, $p<.001$ (Fig.9. – Tool 8).

[Figure 9]

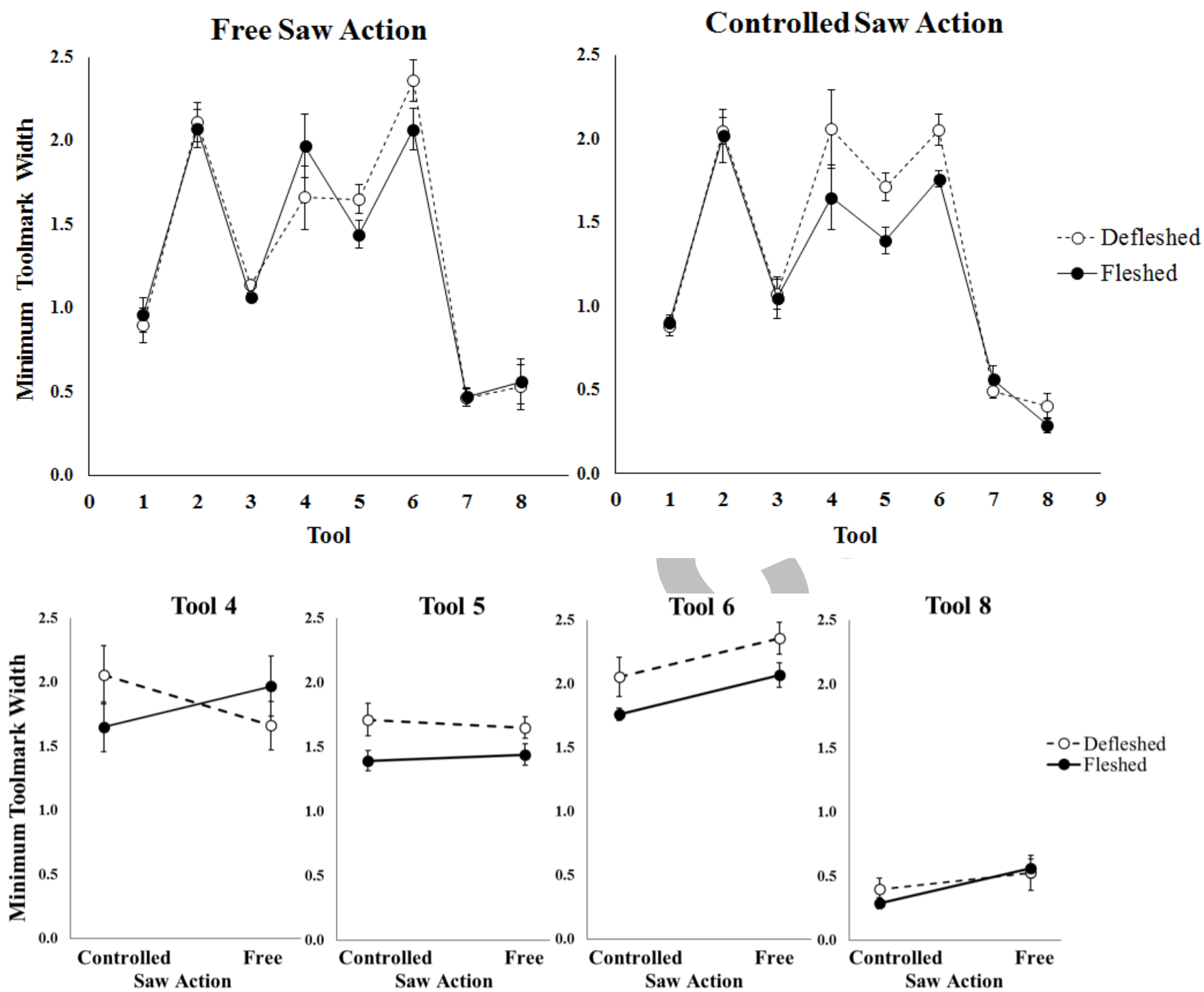

Figure 9. Top – Mean toolmark width and $CI^{95\%}$ for each tool created either on fleshed or defleshed bone and by either a free of controlled saw action. Bottom – Mean toolmark width and $CI^{95\%}$ for tools 4-6 and 8 split by saw action and tissue presence.

DISCUSSION

Toolmark analysis is an important forensic tool particularly in cases of victim dismemberment. Previous studies have explored the use of saw mark properties such as width and shape to determine the tools used to create them [1-18]. In the current study 270 saw marks were created over two experiments using eight different tools. The toolmarks were created on human bone using four different methodological conditions with either fleshed or defleshed bone and with either controlled or free sawing actions. Toolmarks were micro-CT scanned with toolmark properties being categorised or measured by two independent raters. The study aimed to determine whether; i) micro-CT is appropriate for toolmark visualisation, ii) toolmarks differed significantly, iii) these differences could be used to predict the tools responsible, and iv) methodological factors make a significant difference to the toolmarks.

4.1. Micro-CT Imaging

The first research question considered whether micro-CT is an appropriate technology for visualising and measuring qualitative and quantitative toolmark properties and if so, how reliable these measurements are when extracted from micro-CT data. As demonstrated in previous studies, micro-CT imaging permits 3-dimensional non-destructive visualisation, documentation and measurement of toolmark properties. [14-19]. In this study, micro-CT allowed many qualitative toolmark properties to be visualised including wall edge shapes, troughs, scratch marks, tooth hop, bone islands, floor dip, striation marks, exit chipping and removal of cortical bone (Fig.5.) Furthermore, using a standardised process outlined in section 2.4, 2-dimensional shape profiles could be extracted (Fig.6.) and used for the measurement of quantitative toolmark properties. It is unlikely that some of the toolmark properties revealed by micro-CT, such as wall angle and trough angles, would have been possible to measure using other imaging methods [19].

This study followed up on Pelletti et al's recommendation to explore the use and reliability of micro-CT for quantitative toolmark analysis on a large sample [18]. The inter-rater reliability of the toolmark measurements between two rators in this study was high. This suggests that micro-CT, along with the 2-dimensional image extraction process used in this study, allows for reliable saw mark analysis further confirming Pelletti et al's previous findings but using a much larger sample size.

4.2. Toolmark differences

The second question was whether toolmark properties differ statistically when created with different tools. In both Experiment 1 and 2, all quantitative toolmark properties created by each of the eight tools significantly differed from each other with the exception of a few tool pairs as noted in post hoc analyses - this can be seen graphically (Fig. 7.). This not only confirms the findings from previous studies [8, 10-12] but also expands on existing work through the addition of toolmark properties not yet explored in the literature i.e. wall and trough angles. The authors also noted that the variation the 30 toolmarks created for each tool was low for both quantitative and qualitative measures, particularly in Experiment 2.

For categorical properties in Experiment 1, profile shapes for each tool were consistent: Tool 1 was 100% square; tool 2 was 100% W with 58% being truncated; tool 3 was 89% truncated W; tool 4 was 97% W with 27% being truncated; tool 5 was 100% square; tool 6 was 100% U; tool 7 was 89% Y; and tool 8 was 100% V. Edge shapes were also consistent with the exception of tool 1 where the difference between necked and straight edges was marginal. Although categorical properties were not included in any statistical analysis in this study, previous work has demonstrated their diagnostic capability [7,10,12].

4.3. Tool prediction

The third research question was whether, across all conditions, there is a significant correlation between toolmark width and tool blade width and if so what percentage of toolmark width is able to predict tool blade width. Clearly in this study different tool classes resulted in significantly different toolmarks (Section 3.2) and therefore the ability to statically predict the culprit tool is promising. Given that floor width properties were; i) present and measurable across all profiles including convex shapes, ii) highly reliable between raters, iii) likely to correlate with a tool property, i.e. tool blade thickness, only width was used as the predictor variable in a logistic regression. In order to explore a more conservative and potentially more realistic forensic example, a second experiment was conducted using the same eight tools as Experiment 1. In Experiment 2, five toolmarks were created with each tool on defleshed human tibiae set in ballistic gelatin to replace the absent human flesh. These circumstances might reflect a homicide case where some tools are suspected of having been used during a dismemberment and the forensic examiner wishes to statistically determine which tool created the toolmarks left on the human remains. In this case the examiner could purchase identical saws and use them to create known saw marks in the lab to then construct a statistical model for predicting the unknown tool responsible for the saw marks left on the victim. In this study this approach resulted in a regression model, developed from toolmark data in Experiment 2 which predicted (across all methodological conditions i.e. tissue presence and saw action) 94% of tool blade width in Experiment 1. Although positive, this accuracy drops to 39% when predicting only marks made in the fleshed bone conditions and 19% in the fleshed, free saw action condition which is of course the most ecologically valid condition used in this study.

The regression model only used toolmark width, however, it would be possible to exploit other toolmark properties such as profile shape to increase accuracy through the use of other statistical approaches that allow combinations of toolmark properties such as decision tree [12]. Furthermore, no qualitative properties such as striation details, bone islands, floor dip etc were considered in this study. Although qualitative properties are potentially more subjective and difficult to score, they likely offer further increases in accuracy and therefore further research on this is recommended.

4.4. Methodological factors

The fourth and final research question in this study asked whether toolmarks differ if created under different methodological conditions namely fleshed vs defleshed bone and using either a controlled vs free sawing technique. In order to address numerous questions such as the effect of tool wear or tissue burning, saw mark studies use different experimental methodologies with variations in tissue presences (fleshed and defleshed), bone samples (animal, human and synthetic analogs), sawing action (ranging from highly controlled with a mechanical setup through to free saw actions with human volunteers). Nogueria et al., found significant differences between human and pig femurs [13] suggesting that human substitutes are not suitable replacements for toolmark studies looking to offer ecologically valid results. In the current study 230 toolmarks were created (the largest sample size to date) over four methodological conditions; human bone, fleshed or defleshed, with two human sawing actions, either controlled or free. All other potentially confounding variables such as the human donor, bone, anatomical side, toolmark placement and order were counterbalanced. The results indicate that, despite toolmark differences remaining in all methodological conditions, for some tools both the saw action and whether the bone is fleshed or not can impact the toolmark width (Fig.8.) and that this impact is inconsistent across different tools (Note tools 4-6 and 8 in Fig.9.). Furthermore, despite the good inter-rater reliability of minimum width, agreement was significantly lower in the fleshed tissue with free saw action and all other conditions in Experiment 1 as assessed by the absences of overlap at the 95% confidence intervals. These results suggest that methodological differences between studies can have a significant effect on the very measures they aim to explore as well as the inter-rator reliability. The authors suggest that although the use of fleshed human bone and a free saw action is logically difficult, its use is essential for toolmark research to be applicable for forensic practitioners.

4.5. Limitations and further work

First, this study used two independent raters both using the same images extracted from the 3-dimensional micro-CT data to make the toolmark measurements. Further work should test several raters that both independently extract the images from the micro-CT data before categorizing and measuring the toolmarks to determine whether the reliability remains high under this more conservative approach.

Second, like previous work, this study used a limited number of tools and crucially these tools belonged to different classes. Whether toolmark differences can be found and used to predict culprit tools within tool classes has scarcely been researched [7] despite being essential to allow more specific application in forensic cases. The authors plan to conduct a similar study using a larger sample of tools accounting for between, within and identical saw models with toolmarks imaged again being documented using micro-CT. This would allow the creation of a database for which unknown saws could be compared and further work in the development of statistical methods for tool prediction could be carried out.

Finally, future studies should explore how to quantitatively categorise tool blade properties so that they can be correlated to toolmark properties other than floor width. Examples might include the angle of the tool blade for correlation to the wall angle or the cross-sectional shape of the tool that could be mapped onto the toolmark shape profile. The definition of these additional tool properties for use in statistical models would allow for increased accuracy in tool prediction.

CONCLUSIONS

Using a large sample of saw marks created on human bone this study has further demonstrated micro-CT as a powerful and reliable imaging method for the visualisation and measurement of saw mark properties. Based on 270 saw marks, it was found that eight tools created significantly different toolmark properties on human bone in various methodological conditions. A regression model developed using toolmark width from Experiment 2 predicted 94% of tool widths in Experiment 1. Finally, tissue presence and saw action significantly and inconsistently influenced toolmark properties indicating that experimental methodology is important for ensuring ecologically valid data.

REFERENCES

1. **Bonte**, W. (1975). Tool marks in bones and cartilage. Journal of Forensic Science, 20(2), 315-325.
2. **Andahl**, R.O., (1989) The examination of saw marks. Journal of the Forensic Science Society, 18(12):31-46.
3. **Guilbeau**, M. G. (1989). The analysis of saw marks in bone. Masters Thesis
4. **Symes**, S., Morphology of Saw Marks in Human Bone: Identification of Class Characteristics. PhD thesis, University of Tennessee, 1992.
5. **Rutty**, G. N., & Hainsworth, S. V. (2014). The dismembered body. In Essentials of Autopsy Practice (pp. 59-87). Springer London.
6. **Symes**, E Chapman, C Rainwater, L Cabo, and S Myster. Knife and Saw Toolmark Analysis in Bone: A Manual Designed for the Examination of Criminal Mutilation and Dismemberment. NCJRS Publication No.232864, Rockville. MD, 2010.
7. **Saville**, P.A., S.V. Hainsworth, and G.N. Rutty. Cutting crime: the analysis of the uniqueness of saw marks on bone. International Journal of Legal Medicine, 121(5):349-357, 2007.
8. **Marciniak**, S. M. (2009). A preliminary assessment of the identification of saw marks on burned bone. Journal of forensic Sciences, 54(4), 779-785.
9. **Freas**, L. E. (2010). Assessment of Wear-Related Features of the Kerf Wall from Saw Marks in Bone. Journal of forensic Sciences, 55(6), 1561-1569.
10. **Capuani**, C., Guilbeau-Frugier, C., Delisle, M. B., Rougé, D., & Telmon, N. (2014). Epifluorescence analysis of hacksaw marks on bone: Highlighting unique individual characteristics. Forensic science international, 241, 195-202.

11. **Bailey,** J. A., Wang, Y., Van De Goot, F. R., & Gerretsen, R. R. (2011). Statistical analysis of kerf mark measurements in bone. Forensic science, medicine, and pathology, 7(1), 53-62.

12. **Love**, J. C., Derrick, S. M., Wiersema, J. M., & Peters, C. (2015). Microscopic saw mark analysis: an empirical approach. Journal of forensic sciences, 60(s1).

13. **Nogueira**, L., Quatrehomme, G., Rallon, C., Adalian, P., & Alunni, V. (2016). Saw marks in bones: A study of 170 experimental false start lesions. Forensic science international, 268, 123-130.

14. **Thali**, M.J., U Taubenreuther, M Karolczak, M Braun, WBrueschweiler, WA Kalender, and R Dirnhofer. Forensic microradiology: Micro-computed tomography (micro-ct) and analysis of patterned injuries inside of bone. J Forensic Sci., 48(6):1336-42, 2003.

15. **Rutty**, G. N., Brough, A., Biggs, M. J. P., Robinson, C., Lawes, S. D. A., & Hainsworth, S. V. (2013). The role of micro-computed tomography in forensic investigations. Forensic science international, 225(1), 60-66.

16. **Baier**, W., Norman, D. G., Warnett, J. M., Payne, M., Harrison, N. P., Hunt, N. C., Burnett, B. A. and Williams, M.A. (2017). Novel application of three-dimensional technologies in a case of dismemberment. Forensic science international, 270, 139-145.

17. **Baier**, W., Warnett, J. M., Payne, M., & Williams, M. A. (2018). Introducing 3D printed models as demonstrative evidence at criminal trials. Journal of forensic sciences, 63(4), 1298-1302.

18. **Pelletti**, G., Viel, G., Fais, P., Viero, A., Visentin, S., Miotto, D., ... & Giraudo, C. (2017a). Micro-computed tomography of false starts produced on bone by different hand-saws. Legal Medicine, 26, 1-5.

19. **Pelletti**, G., Cecchetto, G., Viero, A., Fais, P., Weber, M., Miotto, D., ... & Giraudo, C. (2017b). Accuracy, precision and inter-rater reliability of micro-CT analysis of false starts on bones. A preliminary validation study. Legal Medicine, 29, 38-43.

20. **Norman**, D. G., Watson, D. G., Burnett, B., Fenne, P. M., & Williams, M. A. (2018). The cutting edge—Micro-CT for quantitative toolmark analysis of sharp force trauma to bone. Forensic science international, 283, 156-172.

21. **Konopka**, T., Strona, M., Bolechała, F., & Kunz, J. (2007). Corpse dismemberment in the material collected by the Department of Forensic Medicine, Cracow, Poland. Legal Medicine, 9(1), 1-13

22. **Porta**, D., Amadasi, A., Cappella, A., Mazzarelli, D., Magli, F., Gibelli, D., ... & Cattaneo, C. (2016). Dismemberment and disarticulation: a forensic anthropological approach. Journal of forensic and legal medicine, 38, 50-57.

23. **Norman**, D., Metcalfe, A. J., Barlow, T., Hutchinson, C. E., Thompson, P. J., Spalding, T. J., & Williams, M. A. (2017). Cortical Bony Thickening of the Lateral Intercondylar Wall: The Functional Attachment of the Anterior Cruciate Ligament. The American journal of sports medicine, 45(2), 394-402.

24. **Norman**, D. G., Getgood, A., Thornby, J., Bird, J., Turley, G. A., Spalding, T., & Williams, M. A. (2014). Quantitative topographic anatomy of the femoral ACL footprint: a micro-CT analysis. Medical & biological engineering & computing, 52(11), 985-995.

25. **Fleiss**, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5), 378.

26. **Koo**, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine, 15(2), 155-163.

SUPPLEMENTARY MATERIALS

Table 1.
Randomised Toolmark positions for each bone sample

Tissue	Bone	Side	Saw Action	Distal End First								Proximal End First							
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Defleshed	Femur	Anterior	Free	5	7	8	4	2	3	6	1	5	6	7[c]	1	8	3	4	2
			Controlled	8	1	7	3	5	4	2	6	7	2	3	1	4[a]	6	8	5
		Posterior	Free	8	4	5	2	1	3	6	7	3[c]	5	8	7[b]	6	1	4	2
			Controlled	5	4	1	2	3	6	8	7	8[a]	4	3	6	1	5	7	2
	Tibia	Anterior	Free	8	4[a]	7[a]	6	5	3	2	1	1	6	7	4	3	2	5	8
			Controlled	1	2	7	3	5	6	4	8	1	6	3	8	2	5	7	4
		Posterior	Free	3	8	4	7	6	5	2	1	8	7	5	6	2	3[b]	4[a]	1[a]
			Controlled	1	2	7	6	3	8	4	5	7	3	2	4	1	6	8	5
Fleshed	Femur	Anterior	Free	4	5	2	7	1	3	8	6	6	7	2	5[b]	1	3	4	8[a]
			Controlled	4	2	7	5	6	8	1	3	5	1[b]	6	2	8	3	7	4
		Posterior	Free	3	5	1	2	4	8	6	7	2	8	7	3	1	5	4	6
			Controlled	6	5	2	3	8	7	4	1	1	8	2	4	6	5	7	3
	Tibia	Anterior	Free	3	4	6	8	5	1	2	7	2	7	6	4	3	1	5	8
			Controlled	8	2	5	1	6	4	3	7	3[a]	6	4	2	5	8	1	7
		Posterior	Free	2	7	5	3	4	8	6	1	6	8	3	1	7	5	2	4
			Controlled	6[a]	1	5	7	3	8	4	2	5	6	8	1	4	7[c]	2[c]	3[c]

a: Toolmark either not present, visible or only resulted in faint scratches
b: No false start toolmark due to a complete sectioning of the sample
c: Toolmark located between two scan points

Table. 2
Inter-rator Agreement, Cohen's Kappa and Intraclass correlation coefficient, for each experiment and each independent condition in Experiment 2.

	Experiment 1					Experiment 2
	Fleshed		Defleshed		Combine	Semi Fleshed
	Controlled	Free	Controlled	Free		Controlled
Profile Shape	κ = .76, [d] $CI^{95\%}$ .64-.87	κ = .72, [d] $CI^{95\%}$ .60-.83	κ = .70, [d] $CI^{95\%}$ .57-.82	κ = .70, [d] $CI^{95\%}$ .58-.83	κ = .72, [d] $CI^{95\%}$ .66-.78	κ = .77, [d] $CI^{95\%}$ .63-.92
Edge Shape	κ = .60, [c] $CI^{95\%}$ .41-.78	κ = .05, * $CI^{95\%}$ 0-.28	κ = .34, [b] $CI^{95\%}$ .11-.56	κ = .43, [c] $CI^{95\%}$ .20-.65	κ = .39, [b] $CI^{95\%}$ .27-.50	κ = .72, [d] $CI^{95\%}$ .53-.90
Minimum Width	ICC = .96, [e] $CI^{95\%}$ .94-.98, F(57)=27.3	ICC = .92, [e] $CI^{95\%}$ .86-.95, F(61)=12.3	ICC = .98, [e] $CI^{95\%}$ .96-.99, F(61)=41.6	ICC = .96, [e] $CI^{95\%}$ .94-.98, F(55)=26.9	ICC = .96, [e] $CI^{95\%}$ .94-.97, F(237)=22.9	ICC = .95, [e] $CI^{95\%}$ .90-.97, F(39)=19.1
Wall Angle	ICC = .92, [e] $CI^{95\%}$ .87-.96, F(57)=13.6	ICC = .85, [d] $CI^{95\%}$ .74-.91, F(61)=6.5	ICC = .92, [e] $CI^{95\%}$ .87-.95, F(60)=13.3	ICC = .90, [e] $CI^{95\%}$ .83-.94, F(55)=10.0	ICC = .90, [e] $CI^{95\%}$ .87-.92, F(237)=9.9	ICC = .92, [e] $CI^{95\%}$ .85-.96, F(38)=12.9
Trough Height	ICC = .98, [e] $CI^{95\%}$ .78-1.0, F(18)=143.7	ICC = .98, [e] $CI^{95\%}$ .94-.99, F(22)=64.1	ICC = .99, [e] $CI^{95\%}$ .82-1.0, F(21)=231.7	ICC = .96, [e] $CI^{95\%}$ .09-.99, F(18)=111.7	ICC = .98, [e] $CI^{95\%}$ .86-.99, F(82)=105.1	ICC = .97, [e] $CI^{95\%}$ .68-1.0, F(14)=1122.1
Trough Shallow Angle	ICC = .66, [d] $CI^{95\%}$ .16-.87, F(18)=3.0	ICC = .70, [d] $CI^{95\%}$ .27-.87, F(22)=3.2	ICC = .36, [a] $CI^{95\%}$ -.59-.74, F(21)=1.5	ICC = .90, [e] $CI^{95\%}$ .75-.96, F(18)=11.0	ICC = .68, [d] $CI^{95\%}$ .50-.79, F(82)=3.1	ICC = .94, [e] $CI^{95\%}$ .82-98, F(14)=15.6
Trough Deep Angle	ICC = .85, [d] $CI^{95\%}$ .61-.94, F(18)=6.5	ICC = .82, [d] $CI^{95\%}$ .57-.92, F(22)=5.3	ICC = .61, [#] $CI^{95\%}$ .06-.84, F(21)=2.5	ICC = .93, [e] $CI^{95\%}$ .79-.97, F(18)=15.6	ICC = .80, [d] $CI^{95\%}$ .69-.87, F(82)=4.8	ICC = .91, [e] $CI^{95\%}$ .74-.97, F(14)=11.3
Combine Quantitative Measures	ICC = .97, [e] $CI^{95\%}$ .96-.97, F(461)=3.12	ICC = .96, [e] $CI^{95\%}$ .95-.96, F(509)=23.5	ICC = .94, [e] $CI^{95\%}$ .93-.95, F(501)=17.6	ICC = .94, [e] $CI^{95\%}$ .92-.95, F(437)=15.8	ICC = .95, [e] $CI^{95\%}$ .95-.96, F(1911)=20.4	ICC = .99, [e] $CI^{95\%}$ .98-.99, F(203)=64.5

- *All p values <0.001 with two exceptions; *p = .581 and [#]p = .019*
- *Approximate agreement levels for Kappa are highlight as: a < 0.20 Poor; b = 0.21-0.40 Fair; c = 0.41-0.60 Moderate; d = 0.61-0.80 Good; and e = 0.81-1.00 Excellent [25]*.
- *Approximate agreement levels for ICC are: a < 0.5 Poor; c = 0.51-0.75 Moderate; d = 0.76-0.9 Good; e = >0.9 Excellent [26]).*
- *For minimum width, agreement was significantly different between Fleshed Tissue with Free Saw Action and all other conditions in Experiment 2 as assessed by the absences of overlap at the 95% confidence interval.*

Table. 3

Experiment 2 - ***Categorical Properties*** *for each tool (data expressed as absolute frequency). N.B. t-W is truncated W shaped.*

	Edge Shapes			Profile Shapes					
Tool	**Straight**	**Necked**	**Alternative**	**Square**	**W**	**t-W**	**U**	**V**	**Y**
1	13	17		28					
2	24	4	3		13	18			
3	28			3		25			
4		1	28	1	22	6			
5	30	1		31					
6	22	9		10			21		
7	28							3	25
8	30							30	

Table. 4

Experiment 2 - Quantitative Toolmark Properties for each tool (Data are expressed as mean ± standard deviation) and a one-way Welch ANOVA with independent variable of tool (1-8) and dependant variable of toolmark property.

Tool No.	**Minimum Width (mm)**	**Maximum Width (mm)**	**Wall Angle (°)**	**Trough Height (mm)**	**Trough Shallow Angle (°)**	**Trough Deep Angle (°)**
1	0.91 ± 0.12	1.03 ± 0.21	4 ± 4	-	-	-
2	2.06 ± 0.18	2.23 ± 0.26	17 ± 16	0.67 ± 0.14	54 ± 10	39 ± 8
3	1.08 ± 0.13	1.23 ± 0.15	18 ± 17	0.14 ± 0.05	66 ± 15	70 ± 18
4	1.84 ± 0.35	2 ± 0.36	9 ± 12	0.71 ± 0.46	55 ± 11	53 ± 11
5	1.55 ± 0.19	1.77 ± 0.23	2 ± 2	-	-	-
6	2.07 ± 0.27	2.26 ± 0.29	27 ± 16	-	-	-
7	0.5 ± 0.09	-	35 ± 11	-	-	-
8	0.44 ± 0.17	-	58 ± 17	-	-	-
Welch ANOVA	$F(7,98)=442$ $p<0.001$	$F(7,97)=425$ $p<0.001$	$F(7,91)=88$ $p<0.001$	$F(2,42)=210$ $p<0.001$	$F(2,50)=6$ $p=0.004$	$F(2,46)=38$ $p<0.001$

Assumptions: *There were no outliers, as assessed by boxplot. With the exception of: wall angle property in tools 2-3 and trough angle deep and shallow for tool 3, data were normally distributed as assessed by Shapiro-Wilk test (p > .05). There was heterogeneity of variances for all properties, as assessed by Levene's test of homogeneity of variances (p < .05). Therefore, a robust Welch ANOVA is reported for all toolmark properties.*

Follow up post-hoc *turkey HSD t-tests reveal that: Minimum and Maximum width differed between all tools (p<0.001) except between tools 2&6, and 7&8; Wall Angle differed between all tools (p<0.005) except tools 1&4, 1&5, 2&3, 2&4, 2&6, 3&4, 3&6, 4&5, 6&7 and 7&8; Trough height and trough angle shallow differed between all appropriate tools (p<0.005) except between tools 2&4; and Trough angle deep differed between all tools (p<0.001).*